American Journal of Medical Genetics (Neuropsychiatric Genetics) 74:286–288 (1997)

Selection of Homogeneous Populations for Genetic Study: The Portugal Genetics of Psychosis Project

Carlos N. Pato,[1,2]* M. Helena Azevedo,[2] Michele T. Pato,[1,2] James L. Kennedy,[1,2,3] Isabel Coelho,[2] Ana Dourado,[2] Antonio Macedo,[2] Jose Valente,[2] Carlos Paz Ferreira,[2] Jose Madeira,[2] Joana Gago da Camara,[2] Margarida Moniz,[2] and Connie Correia[1]

[1]*Department of Psychiatry, SUNY at Buffalo, Buffalo, New York*
[2]*Department of Psychiatry, Faculdade de Medicina, and Center for Neuroscience, Universidade de Coimbra, Coimbra, Portugal*
[3]*Neurogenetics Section, Clarke Institute, University of Toronto, Toronto, Canada*

Molecular genetic studies of psychiatric disorders must face the possibility that despite the significant contribution of genetic factors to the expression of syndromes like schizophrenia, these syndromes may be a heterogeneous collection of genetic and non-genetic illnesses. These illnesses may be etiologically distinct from each other and still share many clinical features in common. Linkage studies of families with multiple affected members tend to favor the selection of genetic forms of a syndrome but can still represent a heterogeneous set of different genetic illnesses. To limit the potential genetic heterogeneity of a study sample, we selected a population that was geographically isolated and was historically relatively genetically homogeneous. We then assessed the relative level of homogeneity utilizing a surname analysis of the population of the Azores, mainland Portugal, rural USA, and urban USA. The average number of families with the same last name corrected for population size in the Azores is 30.88, in Coimbra it is 21.42, compared to 1.13 in a rural American population and 0.38 in an urban American population. The results of this analysis indicate that the Azores have the highest degree of homogeneity, and mainland Portugal has a high degree of homogeneity. Am. J. Med. Genet. 74: 286–288, 1997. © 1997 Wiley-Liss, Inc.

KEY WORDS: surname analysis; genetic homogeneity; psychosis

INTRODUCTION

It is becoming increasingly clear that to find genes for psychiatric disorders we must design and analyze studies in a manner that is suitable for complex diseases. We have designed such a study for psychotic disorders that includes four major components: 1) homogeneity of the sample, 2) converging analytic strategies, 3) careful definition of the phenotype, and 4) collaborative studies. In this paper we describe use of an efficient, low cost method for measuring homogeneity in a population.

The choice of the Portuguese population for this study is based on the hypothesis that the population is characterized by a higher degree of homogeneity than populations found in the United States. We have focused special attention to the population of the Azores. The Azores were settled over 500 years ago almost exclusively by the Portuguese. The islands had no native population when they were first settled by Portugal in the early 1400s. The settlement of the islands was a methodical program that started with populating the major islands with livestock and allowing for a period of over 30 years before the first human settlement in order to provide for the needs of the settlers. Settlement of the islands was also programmed with groups of families being awarded land and the right to settle different areas. This history has been extensively reviewed in a recent book by Costa [1991].

The Azores are a Portuguese state made up of a nine island archipelago in the Atlantic ocean with a population of 252,000. Given the pattern of settlement of these originally deserted islands and the relatively low number of founders, this may be a valuable population in which to detect linkage disequilibrium. The Azores have a centralized health system. All ten psychiatrists on the islands are collaborating with us on this project. The main focus of our study is the Azorean population with the rest of the population derived from continental Portugal. The majority of the Azorean population is derived from this population base. The first phase of our study will focus on families with multiple affected members with schizophrenia, employing both linkage

*Correspondence to: Dr. Carlos N. Pato, Associate Professor and Clinical Director, Department of Psychiatry, Buffalo General Hospital, 80 Goodrich Street, Buffalo, NY, 14203. Email: cmpato@aol.com

Received 6 May 1996; Revised 8 November 1996

TABLE I. Surname Analysis

	USA-urban (1.5 million)	USA-rural (250,000)	Portugal-Coimbra (85,000)	Portugal-Azores (250,000)
Mean (phones/surname)	5.64	2.83	18.21	77.22
Mean (phones/surname) per 100,000 pop.	0.38	1.13	21.42	30.88
Maximum phones/surname	279	303	835	1,766
Surname count/10,000 phones	1,772	3,536	549	129

analysis and non-parametric analytic strategies. We project including approximately 200 families, segregating for the disease. Given the average number of affected relatives in the 60 families identified to date, this includes over 500 affected family members with schizophrenia. The careful diagnostic definition of phenotype is based on detailed structured clinical data employing the Diagnostic Interview for Genetic Studies (DIGS), which we have translated into Portuguese [Azevedo, 1993]. A fundamental feature of the study will be prospective follow-up of patients and relatives participating in the study.

Prospective clinical follow-up is scheduled with each subject to be performed annually after the initial interview. This will allow for confirmation of stability of diagnosis and the identification of newly ill family members. If possible, this prospective aspect of the study will continue beyond the two year point. The Azores may be a unique site for prospective follow-up as all potential treatment sites are integrated and participating in the study. Our program is designed to capture a very complete history of the patient's illness, as well as to be able to follow most subjects prospectively for a long period of time. This will be extremely valuable for achieving diagnostic certainty and minimizing false positives.

We have adopted the diagnostic hierarchy selected by the "NIMH Genetics Initiative" in order to optimize comparability with the families studied by this and other collaborative groups. Our research team consults actively with the collaborative efforts both in the United States (NIMH and VA) and in Europe (EU and ESF). This has led us to maintain our diagnostic methods compatible with both major collaborations and to develop close relationships with a number of collaborating laboratories.

To test the hypothesis that the population is characterized by a higher degree of homogeneity than populations found in the United States, we performed a surname analysis. Surnames have been shown to be excellent highly polymorphic markers for the Y chromosome, given western conventions. Over 1 million surnames exist in western culture [Lasker, 1985]. In the vast majority of cases the surname is passed from father to child. In Portuguese culture it is also common to retain the maternal surname as the next to last name. Our hypothesis was that the higher the degree of homogeneity of the population, the smaller the number of surnames ascertained and the greater the number of phone numbers assigned to the same surname both on average and in terms of total number of phone numbers assigned to a single surname.

METHODS

For our study, we took four populations. The first was the Azores, the second was from the city of Coimbra on mainland Portugal, the third was a rural American population of approximately the same population (250,000), and the fourth was an urban American population of 1.5 million people. The residential phone listings for all four populations were compared showing that all four had approximately 20,000 phone numbers per 100,000 population. Given the comparable level of phone ownership, there was no evidence for a cultural or economic bias. Furthermore the rate of one phone listing per five people is consistent with roughly one phone number per nuclear family.

Surnames were ascertained by including every name present on every tenth page of the phone books for both the populations in Portugal and the rural American population and then counting the number of phone numbers assigned to the same surname. For the urban American population, which was six times larger, we ascertained surnames by including every name present on every fiftieth page of the phone book. If a surname was included, all phone numbers assigned to that surname were counted even if they appeared on previous or subsequent pages.

RESULTS

The results of this analysis indicate that the Azores have the highest degree of homogeneity (see Table I), and mainland Portugal has a high degree of homogeneity. The average number of families with the same last name corrected for population size in the Azores is 30.88, in Coimbra it is 21.42, compared to 1.13 in a rural American population, and 0.38 in an urban American population. In the Azores, in a sample of 18,765 phone numbers only 243 family names were identified. In a sample of 6,778 phone numbers in rural America 2,397 surnames were found. In the Azores, there are 1,766 families with the surname Medeiros, representing the largest number of families with the same last name sampled, while in the USA, the largest sampling was 303 with the surname White in rural USA.

DISCUSSION

These results strongly support the high degree of homogeneity present in the Portuguese population, and even more so in the Azores, and affords us a method to better define the populations in each area. We feel this supports the use of both populations that are likely to be genetically associated, while treating the Azorean population as a special subpopulation that is likely to be even more homogeneous. This method of assessing the relative homogeneity of populations prior to a molecular genetic study has, to our knowledge, not been used before. It is an effective inexpensive method that given western cultural traditions gives us a handle on the genetic make-up of a population.

Clearly, the historical context must also be considered. The specificity of these results rest in part on the known pattern of settlement and patterns of immigration. In Portugal the patterns of immigration are well documented, and the tradition of familial transmission of surnames has been consistent over at least 300 years. However, surname analysis can have serious limitations in the presence of other known historical conventions. For example, the convention common in USA of immigrants adopting modifications of their surnames, or changing surnames, greatly reduces the specificity of the surnames as genetic markers. This known pattern may have at least two types of effects. One effect would be to falsely increase the apparent homogeneity of the population based on the disproportionate adoption of Anglo-American names such as Jones or Smith. The predicted effect on the population in the USA, given the patterns of immigration, would be a reduction in the variety of surnames. Another possible effect would be the potential diversification from a single immigrant surname. For example, the name Correia might be transformed into the names Curry, Curri, Currier, Corea, Chorea, Korea, etc. These two effects may serve to neutralize each other, but the extent to which this happens is unknown. Overall we see a dramatic difference in the degree of homogeneity, demonstrating the marked heterogeneity of the population in the USA. These limitations must be considered in evaluating the results. However, the ability to screen a large population, in a relatively inexpensive fashion, greatly outweighs these limitations.

Genetic studies of the major psychiatric disorders are maturing. It is recognized that the complexities are greater than first estimated. Mitigating against this increase in complexity is the increase in the power and information from the human genome project and the technology used to search for genetic factors in disease. It is most important to recognize that even in complex non-psychiatric disorders, genetic approaches are proving extremely valuable. The technology to scan all the chromosomes (genome scan) has become increasingly comprehensive and efficient. In the past few years several thousand highly informative markers have been developed and are readily available for a PCR-based genome scan for any given genetic disorder. The number of highly polymorphic PCR-based markers discovered on the human genome continues to increase at a rapid rate. It is likely that in the near future the human genome map will have genetic markers spaced at one to two centimorgan intervals. The information power in this set of markers will allow the rapid confirmation or exclusion of any hypothesized genetic linkage finding, especially if the advantages of homogeneous populations are realized. In conclusion, we have described a surname analysis method that greatly facilitates the selection and characterization of populations for molecular genetic studies.

ACKNOWLEDGMENTS

This work was supported by NARSAD Young Investigator Award (C.P.), NARSAD Independent Investigator Award (J.K.), and JNICT grant (M.H.A.,C.P.).

REFERENCES

Azevedo MH, Valente J, Macedo A, Dourado A, Coelho I, Pato MT, Pato CN (1993): "Versao Portuguesa da Entervista Diagnostica para Estudos Geneticos." Psiqui Clin 14(4):213–217.

Costa C (1991): "Etnologia dos Acores." S. Miguel: Lagoa.

Lasker G (1985): "Surnames and Genetic Structure." New York: Cambridge University Press.